CONF-9004177--3

CONF-9004177--3

DE90 017328

ARTIFICIAL INTELLIGENCE FOR SUBJECT INTERVIEWING

Bruce E. Tonn
Richard Goeltz
Lloyd Arrowood
Kerry Hake

Oak Ridge National Laboratory*
Oak Ridge, Tennessee 37831

Advanced Computing for the Social Sciences
Williamsburg, Virginia
April 10-12, 1990

*Operated by
MARTIN MARIETTA ENERGY SYSTEMS, INC.
for the
U.S. DEPARTMENT OF ENERGY
under Contract No. DE-AC05-840R214000

ABSTRACT

This paper has two goals: to discuss, in general terms, issues related to applying artificial intelligence (AI) techniques to computerized interviewing; and to describe two AI-based interviewing systems developed at Oak Ridge National Laboratory. With respect to the former, AI techniques can be used effectively to collect data of complex representation, provide flexibility in collecting data, and improve data validity through real-time reviews. One AI system, ARK, elicits subjects' beliefs on an open-ended range of issues and topics through menu-driven, dialogue-based interactions. The other system, LES, elicits uncertainty assessments related to events, statements and propositions and tailors questions for subjects to explore their uncertainty processing heuristics.

KEY WORDS: artificial intelligence
subject interviewing
knowledge elicitation
computer assisted interviewing
probability elicitation

INTRODUCTION

In some fashion, artificial intelligence can contribute to every aspect of social science. For example, expert systems can provide advice on survey and sample design, help to classify open-ended survey question answers (Appel 1987), and suggest appropriate statistical methods to analyze data (Gale 1986). Machine learning algorithms could be trained to run complex, interactive market simulation programs and could be effective in certain data analysis contexts. This paper has two goals: to discuss, in general terms, issues related to applying artificial intelligence (AI) techniques to computerized interviewing; and to describe two AI-based interviewing systems developed at Oak Ridge National Laboratory (ORNL).

Even within the narrow topic of computerized interviewing, AI has many applications. For example, AI methods are being used to produce intelligent interfaces capable of recognizing voices, parsing natural language messages elicited from subjects, and responding "thoughtfully" to user supplied information. This paper covers methodological issues related to collecting data from subjects using intelligent computer programs. Thus, the central theme of the paper is what intelligent interviewing systems can offer that more traditional data collection approaches cannot.

Distinctions Between AI-Based and Classical Social Science Data Collection Techniques

What characteristics distinguish intelligent interviewing from other social science data collection techniques? We have identified three. The first relates more to a mind-set than to the power of the computer, and it falls under the rubric of data/knowledge representation. Artificial intelligence researchers pay special attention to how data are represented and, in most contexts, the data need to be represented in complex manners. For example, in expert system contexts, data are actually pieces of knowledge that describe a specific part of how the world works. In robotic applications, data represent how real-world activity is put together, and in scheduling applications, the data represent essential details of productions plans and constraints. These data go beyond the basic schemes of value and relationships embodying the sense of meaning, intrinsic value, and mental constraints.

Intelligent interviewing, almost by definition, attempts to collect more complex types of data than would traditional pencil and paper surveys. The data are typically not representable by simple variables such as yes/no, Likert Scale (e.g., 1 through 7), or multiple choice. Instead, the data usually have complex representations and are defined in highly interdependent ways. For example, describing how the world works requires linking at least two concepts, (e.g., an initiating event and a resulting event), with an operator (e.g., causes, as in smoking causes cancer). How the world is put together requires at least two concepts linked by conceptual building blocks. Commonly used building blocks include: IS-A, where one can say that "AIDS IS-A national problem;" and IS-A-PART-OF, where one can say that "the Social Security Administration IS-A-PART-OF the federal government." Such data representations flow from basic artificial intelligence research in knowledge representation (Barr and Feigenbaum 1981). Whereas traditional social science methods could not have collected such data from subjects, an intelligent interviewing approach is predisposed to collect such data.

The second and third distinctions relate to the power of the computer. Second, the computer allows flexibility and uniqueness. Compared to pencil and paper exercises, an intelligent interviewing system can provide subjects with tens, if not hundreds, of response choices. Such systems can also tailor questions based on subject characteristics or based on previous answers. Of course, it is still very much of an art to design experiments where some semblance of order can be distilled from the chaos of allowing each subject to follow his or her own path through the experiment. AI systems can be made flexible by adapting to a subject's interface preferences, instruct subject on system use at appropriate levels, and offer individually tailored multinational messages.

The third distinction relates to rigor, which has two meanings, given our context. On one hand, rigor relates to the validity of data being collected. An intelligent interviewing system can improve data reliability and validity through (1) intelligent checking of subject answers for inconsistencies derived from boredom or misunderstanding of the task at hand, (2) prompting subjects to review their answers, and, if needed, (3) automatic imputation of missing data.

Rigor also relates to describing the transfer of information from subjects into the computer. This is not a problem with closed-form pencil and paper surveys, but it can be a problem with ethnographic interviews, for example. During ethnographic interviews, the interviewer can lead the interviewee in any number of manners, which may not be particularly predictable at the beginning. The protocols from which interviewers operate and the notes or transcripts of actual interviews may be sketchy, wordy, and syntactically complex. It may be extremely difficult and time consuming to translate free-flowing, natural language data into forms amenable to computer representation, analysis, and manipulation. Intelligent interviewing may not yield the same amount of information as an ethnographic interview, nor be as free flowing, but the data will always be rigorously structured and immediately available for computer analysis.

There are other general benefits to computerized interviewing not necessarily specific to intelligent interviewing. First, such instruments treat each subject exactly the same, thereby avoiding possible interviewer bias. Second, since the computer is completely impersonal; computerized interviewing may actually be able to collect information that people would not volunteer to an interviewer, or even commit to paper. Third, if the computer is readily accessible, the computer will be ready at any time convenient for the subject.

Intelligent interviewing is not without problems. For example, in response to the last point, computers are not readily accessible and arranging for an appropriate sample of subjects who are willing and/or able to use computers may be a problem. More daunting problems are the substantial, up-front costs associated with software development and uncertainties inherent in breaking new ground in social science research and, often times, in artificial intelligence. A last problem involves ensuring the confidentiality of the data files both from other users and from malfeasant hackers.

In summary, intelligent interviewing systems can be characterized with respect to complex data, flexibility, and rigor. However, the problems mentioned above must be overcome or at least competently managed before such systems will become widespread. The next section presents two intelligent interviewing systems that highlight the points made in this section.

ARK - Acquiring and Reasoning About Knowledge

Oak Ridge National Laboratory conducts research into the development of intelligent interviewing systems. The first system to be discussed is known as ARK. It was originally developed as an automated knowledge acquisition system to aid in the creation of expert systems (Tonn and Arrowood 1987). However, because of its generality and menu-driven, man-machine dialogue approach, it has been found to be very useful for interviewing applications. ARK is written in Common Lisp for transportability and has been successfully ported to various computers.

ARK's goal is to elicit from subjects their beliefs about a particular subject. Each belief can be characterized by eight pieces of information:

1) agent(s) of change (e.g., AIDS)
2) consequence of the agent(s) (e.g., illness)
3) a propositional primitive describing the relationship between the agent(s) and consequence (e.g., AIDS 'directly causes' illness)
4) likelihood of the consequence given the agent(s) (e.g., AIDS directly causes illness, probability 0.95)
5) sources of belief about the proposition (e.g., news media)
6) salience of the consequence (e.g., illness is an 80 on a 0-100 scale)
7) time of consequence occurrance (e.g., within ten years)
8) confidence in the stated belief (on a 0-100 scale).

In line with the topology developed in the previous section, ARK collects complex data. ARK is also flexible. Subjects can choose from among five propositional primitives (directly causes, prevents, accompanies, associates with, and possibly relates to), three likelihood modalities (probability, percent of time, chances), and eight sources of belief (personal experiences, experiences of acquaintances, the news media, scientific fact, religious doctrine, common knowledge, authoritarian doctrine, and scientific research). In addition, ARK allows subjects to specify any agents of change and any consequences. And, when one belief is completed, ARK will allow subjects to discuss the consequences of consequences of consequences, etc. As a result, ARK is able to tailor itself to what subjects are interested in pursuing. ARK is rigorous in that all the data it collects from subjects is immediately available for analysis. ARK also has the ability to identify logical inconsistencies among a subject's beliefs.

Figure 1 illustrates the kind of data collected by ARK. The figure represents the aggregation of consequences of AIDS, as elicited from 129 University of Tennessee-Knoxville students during sessions with ARK. From left to right, the concept net moves from highly abstract to less abstract consequences. Thus, events are more abstract than events which are physical-biological. Each of the student answers was classified in one of the boxes found at the end of the paths in the tree. For example, 20 consequences are classified in the box labeled "sociological," which is traceable, from left to right, by following the "object" box, then the "abstract" box, then the "societal" box. Eighty-seven consequences are classified in the box "death," which is reached by following the "event" box to the "physical" box to the "biological" box to the "human" box. Table 1 presents examples of actual student responses and how they were classified.

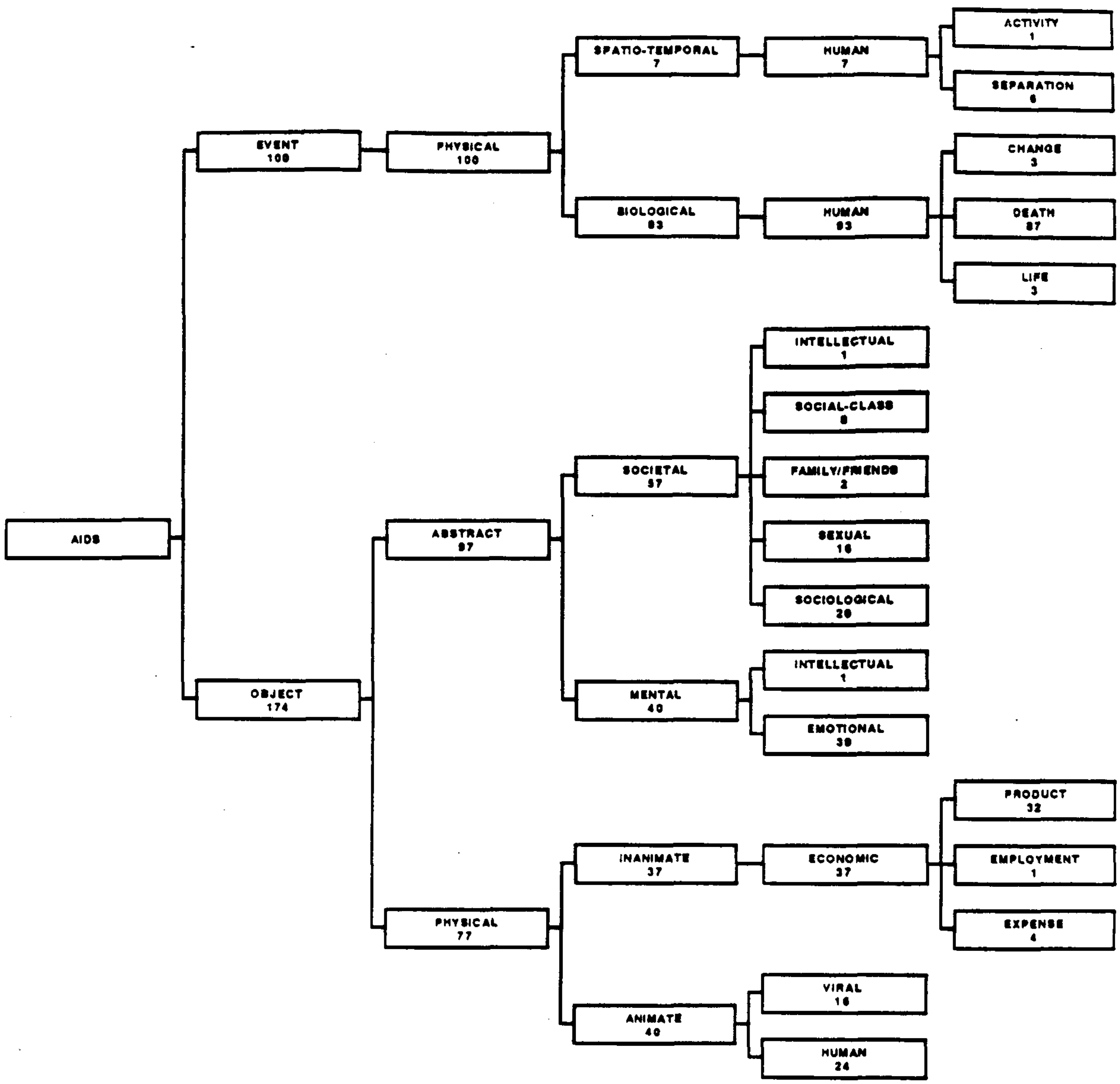

Figure 1. Direct Consequences of AIDS*

*Numbers in Boxes are frequencies.

Table 1. Instances of Frequency Mentioned
Primary Consequences

Class	Consequence	Frequency
Death Human	Individual Death	89
	Population Death	6
Death Non-Human	Death of Animals	6
	Death of Fish	4
Economic Product	Advances in Medical Research	15
	No Cure for AIDS	8
	Research	6
Emotional Abstraction	Fear of AIDS	5
	Fear of Unknown	5
	Disgrace	4
	Fear of Blood Transfusions	4
Environmental Event	Contamination	14
	Destruction of Environment	11
	Destruction of Life	5
Environmental Object	Pollution	27
	Water Pollution	15
	Air Pollution	9
	Hazard to Life	7
	Filth	5
Physical Condition	Ill Health	21
Social Class	Innocent Victim	6
	Social Outcast	5
Sociological Abstraction	Suspicion Sexual	7
	Social Problems	5
Viral Object	Disease	12
	Cancer	10
	Contagiousness	8

As the figure illustrates, ARK data are complex. Combined with the data on choice of propositional primitives, likelihood estimation, source of belief and salience, ARK yielded a unique data set that has contributed to the field of risk analysis (Tonn et al. 1989). Continued research with ARK will require further innovations in data analysis and presentation methods. For example, it is an open question as to how to present in a figure the consequences of the consequences of the consequences, etc. It is also a challenging statistical problem to analyze aggregations of individual belief structures. With an improved interface and increased functionality, ARK holds promise of contributing to advanced computing in the social sciences.

LES - Likelihood Elicitation System

LES elicits from people uncertainty assessments associated with single events (e.g., going to the grocery store on an average day), simple cause and effect statements (e.g., exposure to asbestos resulting in cancer), and simple propositions (e.g., a person who is tall having a GOOD personality). LES also elicits uncertainty assessments associated with compound events, chains of cause and effect statements, and complex propositions. The data are valuable in evaluating the quality of subjects uncertainty assessments and in exploring heuristics subjects use to process uncertainties.

LES accomplishes this task in two sessions. Session 1 elicits simple uncertainty assessments and Session 2 elicits the complex assessments. As one aspect of flexibility, LES constructs Session 2 questions based on the magnitude of the answers for Session 1 in order to find out how subjects combine uncertainty estimates that vary by magnitude. This tailoring has resulted in data that clearly indicate that subjects heuristics change as the magnitudes of the uncertainties being combined change (Tonn and Goeltz 1988).[1]

LES exhibits flexibility in other ways, too. It can ask questions about any topic; daily events, cancer causes and effects, personality traits, and world events have been tried so far. LES also offers a comprehensive set of uncertainty assessment modalities, including probability, chances, percent of time, certainty factors, possibilities, and natural language terms. With respect to the latter, LES allows subjects to construct 324 different terms. Table 2 presents the terms most frequently used by subjects answering questions about cancer and daily events (Tonn et al. 1988).

Several aspects related to reliability have been programmed into LES. First, at the end of Session 1, LES presents to subjects special screens that contain all their answers to the Session 1 questions. The subjects are encouraged to review their answers and LES provides the capability to change one or all previous answers. At the end of Session 2, LES searches for inconsistencies among the answers. For example, if one subject provided the probability of 0.65 for the combination of two events with probabilities 0.80 and 0.90, then this answer would be inconsistent with the probability of 0.75 given for the combination of two events with probabilities 0.70 and 0.60. When such inconsistencies are found, LES presents to subjects the inconsistent answer and asks whether the subject wishes to change the answer. If the subject declines, then LES presents the answers to the single likelihood questions that compose the compound question. The subject can change these answers and then answer additional compound questions or leave all answers as they are.

---

[1]Specifically, data show that whatever heuristics are being used they less and less resemble the product rule as the uncertainties to be combined decrease in magnitude.

Table 2. Most Frequently Used Natural Language Terms by Experiment (% of Answers)

TERM	CANCER EXPERIMENT (N=3229) % OF ANS.	FREQ.RANK	DAILY EVENT EXPERIMENT (N=1005) % OF ANS.	FREQ. RANK
Possible	18.0	1	10.3	2
Unlikely	10.9	2	11.0	1
Impossible	7.5	3	1.5	20
Likely	6.4	4	6.8	4
Uncertain	5.9	5	0.5	32
Highly Unlikely	4.1	6	5.9	5
Very Likely	4.1	7	7.4	3
Very Possible	3.4	8	3.4	8
Very Unlikely	3.3	9	4.6	6
Somewhat Likely	2.6	10	3.4	8
Highly Likely	2.1	11	0.8	24
Somewhat Possible	2.1	11	4.1	7
Certain	2.0	13	2.4	12
Virtually Impossible	2.0	13	0.4	36
Rather Unlikely	1.6	15	1.2	21
Fairly Likely	1.4	16	2.0	15
Extremely Unlikely	1.4	17	2.1	13
Highly Possible	1.3	18	1.2	21
Fairly Possible	1.2	19	1.7	17
Almost Impossible	1.1	20	0.7	26
Somewhat Unlikely	1.0	21	3.4	8
Fairly Unlikely	0.8	25	2.7	11
Extremely Likely	0.8	26	1.6	18
Very Certain	0.7	28	1.9	16
Almost Certain	0.5	30	1.6	18
Extremely Certain	0.3	45	2.1	13

The data structures elicited from subjects by LES are not as complex as those elicited by ARK. However, the means of data capture possess more depth. For example, LES keeps data on all edits made by the subjects. Thus, for example, we have determined that most subjects change answers when LES queries them about answer inconsistency. Also, in the natural language applications where subjects are requested to order from least likely to most likely all the terms they used, LES collects data on subjects sorting commands. This latter data have proved a challenge to analyze.

In summary, LES has proved successful in collecting data unique in the analysis of human reasoning under uncertainty. Its ability to tailor questions and elicit reliable data have been the keys to success. The next version of LES will attempt to elicit nonadditive probabilities, such as those related to belief functions (Shafer 1976) and will have experiments that explore the relationships between time and uncertainty assessment. In the future, it may be possible to provide LES with machine learning abilities that will allow it to hypothesize about a subject's heuristics and then develop questions to support or reject hypotheses.

DISCUSSION

Intelligent interviewing systems offer the potential to advance the state of the art in social science computing. Both ARK and LES have passed initial field testing and have yielded valuable data. Each, in its own way, collects complex data, offers subjects flexiblity, and attempts to improve the reliability of data being collected. Work on these systems has instantiated problems related to up-front development costs and computer accessibility. Computer security was a potential problem when it was found that subjects attempted to access the software at a later point in time. Fortunately, changes in experimental procedures have eliminated this problem.

Continued development of these systems faces numerous challenges. One of the most daunting is the choice of delivery systems. Currently, development work is done on a Symbolics Lisp machine and the software is field tested on VAX equipment. However, there might be benefits to having the software run on IBM PCs or on MacIntoshes or on other general purpose computer workstations. The project team is firmly committed to Common Lisp, but there are choices involved with computer interface and graphics software. We have found that advanced software development environments and advanced AI techniques are not extremely compatible with most environments conducive for computerized interviewing. We have committed to the former with hopes that the latter will meet our technology needs in the future.

With respect to the continued development of both ARK and LES, there is much to be done. ARK could benefit from natural language processing capabilities and more intelligent means of sorting through risk beliefs to find points of interest to discuss with subjects. LES could benefit from an interactive machine learning algorithm, which would guide Session 2 questioning in highly intelligent manners.

REFERENCES

Appel, M. 1987. "Automated Industry and Occupation Coding," Presented at Statistical Tools Seminar on Development of Statistical Expert Systems, Luxemburg, December 1-3.

Barr, A and Feigenbaum, E. 1981. The Handbook of Artificial Intelligence. Vol I, William Kaufman, Los Altos, CA.

Gale, W. 1986. Artificial Intelligence and Statistics. Addision Wesley, Reading, Mass.

Shafer, G. 1976. A Mathematical Theory of Evidence. Princeton University Press, Princeton, NJ.

Tonn, B. and Arrowood, L. 1987. "ARK- Acquiring and Reasoning about Knowledge", Proceedings of the Third Annual Expert Systems Government Conference, Washington, D.C., Oct. 21-23.

Tonn, B. and Goeltz, R. 1988. "People Versus Probability," ORNL/TM-10898, Oak Ridge National Laboratory, Oak Ridge, Tenn.

Tonn, B., Goeltz, R., and Travis, C. 1988. "Elicitation of Natural Language Representations of Uncertainty Using Computer Technology," Proceedings of the Society for Risk Analysis, Washington, D.C., Oct 31-Nov. 2.

Tonn, B., Travis, C., Goeltz, R., and Phillippi, R. 1989. "Knowledge-Based Representation of Risk Beliefs," Risk Analysis, forthcoming.